

The Human Interface Group

“Sounds Good to Me”

An investigation of Spatially Localized Sound for the Personal Computer

The Leonardo Concept System Team

Investigation by:
Lewis Knapp
David Torgersen
Penny Bauersfeld
Michael Chen
Laurie Vertelney
Stephanie Houde
Richard Lyon
“The Users”

Prepared by: Lewis Knapp

The Human Interface Group/ATG
Apple Computer, Inc.
December 31, 1990

Technical Report: 90-19

Table of Contents

Background.....	3
The Plan.....	4
Hardware.....	5
Software.....	8
Experiments.....	8
Conclusions.....	12
Appendix A: The first experiment	17
Appendix B: The “Sonic Configurator” User Test.....	20
References.....	24

Background

Since the inception of the Leonardo Concept System there has been interest in investigating what 3D or Spatially Localized Sound technology would bring to the person who uses a computer to design in three dimensional environments. This report documents work which paralleled an investigation of interfaces for Space Planning in 3D by the Leonardo team during the summer of 1990. (See also ATG HIG technical reports "Space Planning Project Technical Report", "Hands On" 3D Direct Manipulation: The Use of Hands as Interface Elements in a 3D Space Planning Prototype, "Architect at Work Technical Report" and "MacApp Space Planning Technical Report".) Much of the work documented here does not directly address the use of 3D sound in design. The investigation included starting from knowing little about the sound tools and technology available and learning the strengths, weaknesses and applicability of different pieces as we went along. Proposing scenarios of how the things we were working on might be used and showing sketchy prototypes of them to people who were willing to listen was an important part of the work. The audience was a source of valuable ideas which the investigators would have taken a *long* time, if ever, to come up with. They provided a good reality check.

Informally, "3D sound" can be defined as the result of a technology which attempts to place the apparent source of sounds in space so that, for example, recordings of several voices could be played such that the blindfolded listener could accurately point out the positions in the room of the persons speaking. Ideally, the recorded voices would be indistinguishable from the voices of persons in the room, both in position and in audio quality. The effect would ideally be the same as being there.

Research in how this spatiality can be achieved has been done in both research labs and in the audio entertainment industry. Our initial awareness of the technology began with the work which was sponsored locally at the NASA Ames Research Center, where it was developed with the pilot in mind. Today's audio systems capture sounds in a discrete number of channels with the unfortunate byproduct being that the exact location of a sound in space cannot be recreated by the facilities which allow the "unaided" person to accurately locate the sounds about him. The research at Ames focuses on what mechanisms the human uses to localize the sounds so that the spatial dimension can be simulated to organize and hopefully increase the bandwidth of the audio perceptual channel. This research is published and technology has been transferred to the public sector in several ways, most notably by the availability of a device called the "Convolutron" from Crystal River Engineering, Inc.

Other efforts in using spatial sound in the audio entertainment field include "Holophonic" sound and "Q-sound". These techniques seem to be closely guarded forms of black art which are zealously guarded by the owners, who dream of vast sums of cash to unlock the secrets, rather than science or engineering.

The physical means by which the 3D sound illusion is ostensibly created range from inexpensive home speakers to stereo headphones to the full "virtual reality suit" consisting of stereo audio, two channel video helmet, dataglove and headtracker. The advantages of stereo headphones and headtracking combined allow the wearer to reposition the head to in effect induce controlled motion to the sound source and thus resolve ambiguities in position (e.g. front-back confusion) by natural head movements. In our investigation, we considered the cost of the headtracker and chose to see what we could accomplish using only stereo headphones as a base configuration, considering the fact that the Walkman has currently brought headphones into vogue as a source of "personal music". The low cost configuration was in keeping with the Leonardo theme of "3D for everyone".

Finally, who might be the users of 3D sound? We envision providing it as an orientation cue to people who might use computers to design real spaces with familiar acoustic properties, such as architects. We imagine 3D sound effects which range from contrived effects to add conceptual richness to a customers view of a proposed design to a realistic acoustic simulation of the actual reflections of a concert hall or the effects of sound dampening material in an office.

The Plan

Ideally, localized sound would contribute at least another dimension of realism to the interface of a program for moving objects in a three dimensional space. With sound, the synergistic effects would make the combined effects very convincing. The plan at the outset was to understand the use of the sound technology in parallel with the development of other interface elements and integrate it with these interface techniques to produce a demonstration which would enable others to experience 3D sound in a task-oriented setting.

The investigation hinged on temporary use of a Convolvotron, the only available machine currently capable of doing real-time multi-channel sound localization. We used the Convolvotron to add localization of both interaction sounds and environmental sounds to the space planning demo. Interaction sounds are sounds used to reinforce the notion of causality such as playing the sound of a chair sliding when it is moved along the floor. Environmental sounds are background sounds which would help in orienting a user in a setting, such as the sound from a stereo or the voice of a person. Environmental sounds would provide additional cues to supplement what is lost to our peripheral vision when we use a 13" screen to simulate our field of view.

The scenario assumed that the video monitor acted as the viewplane of the user, so the operators eyes (and ears) were in effect fixed to the monitor. Position and view changes only moved what appeared on the screen and in the auditory field. This

assumption, that the head and monitor were "attached" would let us get started without requiring a device for head position tracking.

We made the assumption that we would use high quality sound rather than sound which has known audible artifacts so as to not bias subjective evaluations of the localizing technology by using sounds which just didn't sound good on their own.

Finally, we wanted to use the Convolvotron hardware to master a demo or animation which could be played back later via stereo headphones which would allow others to hear localized sound and judge for themselves whether it "works" for them. The restricted availability of this equipment only makes it possible to hear the effect during public talks and demos, when there are many people waiting to listen as well. Consequently, people are forced to decide whether the effect works for them and whether they like it based on a very short exposure to the medium. The demonstrations are usually highly biased toward presenting the technology in as favorable a manner as possible. With these goals in mind, we set off to assemble the people and pieces required to make it happen.

Hardware

Because the interactive Space Planning prototype we were developing already used all the horsepower we could get to do the 3D display and interaction without sound, it seemed reasonable to use other machines to provide the sound support. A search within ATG for equipment which was available for this experiment led to Rob Wolff's sound laboratory, a facility which is used to produce audio to support scientific visualization video presentations. Rob graciously gave us full access to facilities which we didn't understand at first, but which turned out to be just what we needed. This facility led us quickly to understand the world of the MIDI interface standard and the wide variety of devices for producing music for entertainment.

Briefly, MIDI is a standard which is based on the notion of capturing the gesture associated with musical performance on a keyboard instrument. It has underlying commands such as "note on", "note off" and other parameters which can be used to convey expressiveness, like how hard the key was pressed. A synthesizing keyboard can, for instance, produce audio output and MIDI output as well, so that external MIDI equipment can be used to provide the sounds of additional instruments. The MIDI standard is the "glue" which has allowed multiple vendors of equipment to provide interfaces like keyboards, drum boxes, etc. to be connected to sources of sound in a free, "mix and match" way by musicians who are not necessarily computer literate, with a high degree of acceptance.

MIDI seemed a natural way to cue interaction sounds and, if you generalize the notion of "note" to include recordings of voice clips and background sounds, to cue

the environmental sounds as well. The Macintosh is easily integrated into the MIDI environment via a small, inexpensive box which attaches to one of the serial ports. As we went further down the commercial audio path, other devices which are useful appeared.

Samplers: In order to capture and edit sounds for interaction, we used a Casio FZ 10M sampler, a device which can basically sample sounds to be stored digitally, edit various aspects of the sound (like transposing it so it has different apparent pitches or changing attack and decay envelopes) or just play it on (MIDI) command. It also comes with a large library of existing sounds, some musical, some just plain strange. We also used an E-Mu Systems EMAX 2 in early stages of the project.

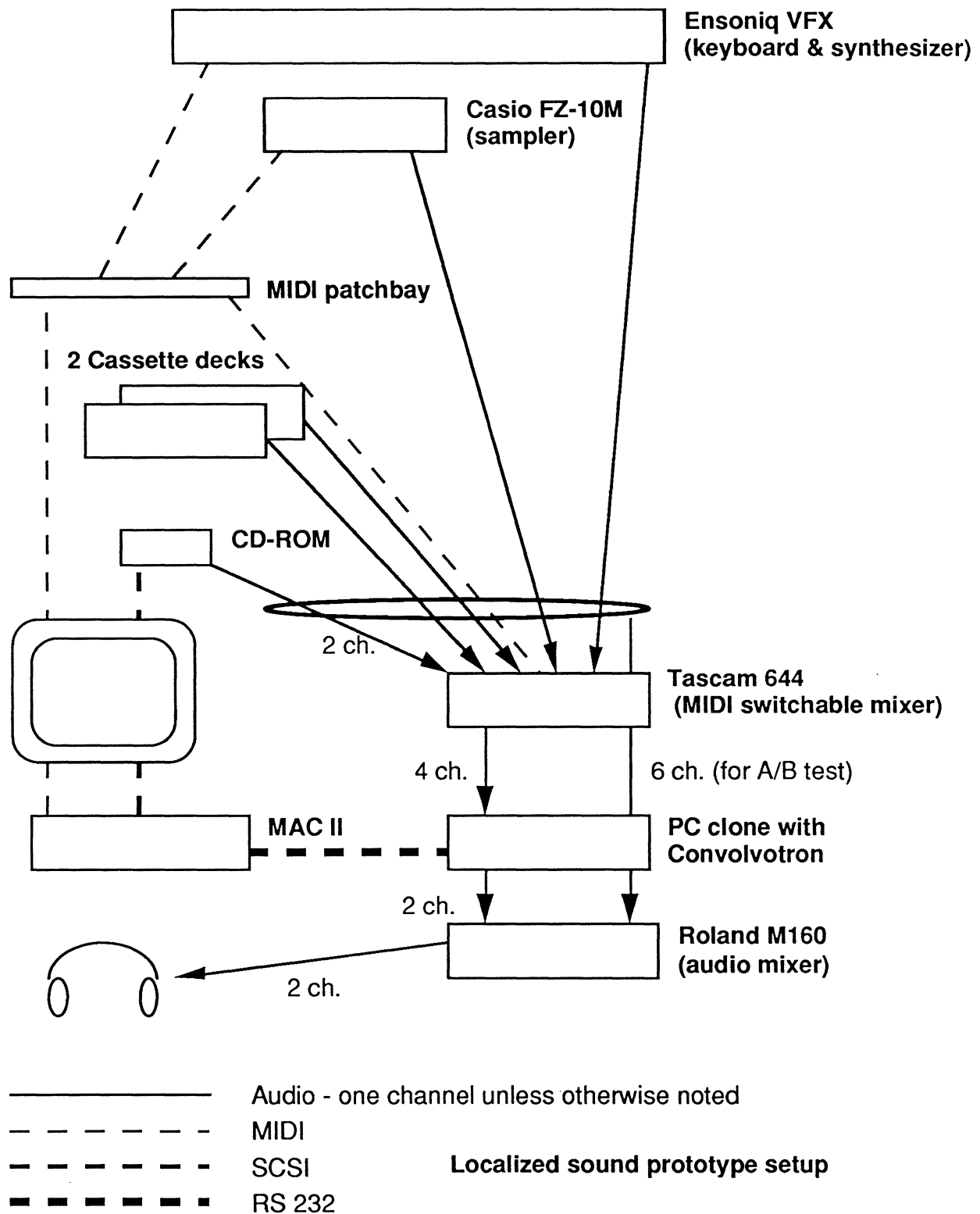
Synthesizers: We also used an Ensoniq VFX synthesizer to generate test tones. Again, a remarkable digital box which allows selection from a large variety of base "voices", combination of these voices, introduction of special effects like reverb and waveform modification in a variety of ways. We used about .01% of its repertoire.

Sound Accelerators/digitizers: A DigiDesign Sound Accelerator was used to digitize pieces of the "Holophonic" sounds from a distributed (copywrited) CD for later analysis of the effect.

Audio devices: An Apple CD-ROM drive was used as a cueable sound source. Cassette decks which were manually started were used to simulate analog sound information channels.

Plumbing: Two audio mixers were used to combine and compare sound sources and provide programmable switching among them. The Roland M 160 had 16 input channels and was used to drive the operator's stereo headphones from a variety of sources. This permitted A/B listening comparison of audio quality with and without the Convolvotron and manual creation of side-to-side effects using the stereo "pan" controls without the localizer. The TASCAM 644 has 8 input and four output channels. It was used to adjust levels going into the Convolvotron so that maximum signal to noise ratios were maintained and to switch between input sound sources via MIDI control.

Convolvotron: The sound localizer consisted of a two board set and a host IBM PC clone. One board was an off-the-shelf Texas Instruments 320/C25 processor development system which had 50 KHz. A/D and D/A converters for audio in and out. It controlled the other board, a custom multiply/accumulate engine with 3 additional A/D and one D/A converter. The pair is sufficient to locate 4 distinct audio sources and drive stereo output channels in real time. The IBM PC is connected to a Mac II which controls sound location parameters via RS232 serial line and runs the interface software.



Software

The Macintosh MIDI manager provided a clean interface to cue sounds from interactive demos like the integrated MacApp space planning prototype. It is straightforward to install and call from a C environment.

HyperCard + Serial Toolkit + MIDI was the combination used to experiment with the Convolvotron and administer tests to the subjects. It allowed us to build a control panel for the Convolvotron quickly and generate sounds via MIDI very easily. It was a convenient way of generating and randomizing test vectors and it made administration of an ever-changing user test simple because the code which administered the test was preserved with the subjects data. David Torgersen, a contractor, provided sound editing and HyperCard programming for much of this effort.

MacroMind Director 2.0 was used extensively for prototyping of space planning interaction tasks and the ready access to MIDI through it made it very straightforward to cue interaction sounds from a Director script. Penny Bauersfeld added sliding and rotating sound cues to a prototype which allowed a user to position a chair in a room.

Michael Chen's MacApp space planning prototype was augmented with localized sound and used the serial interface to the PC to position sounds relative to the observer who was moving furniture and changing his view of the screen. Interaction sounds were triggered by calls to the MIDI toolbox.

Sounds: An early search for sounds from the real world turned up a set of 50 audio CD's called Sound Ideas in the Engineering Library. This is well cataloged and readily available if what you need is a sound grabbed from real life situations. These weren't used, but they are convenient if you need some sounds to digitize.

Experiments

The first phase was to hook the Convolvotron up and try it out. This involved writing the Mac end of a program which talked to vendor-supplied software on the PC which would, for each of four sound sources, locate the source in the user auditory space according to angles of elevation and azimuth and set the volume of the channel. Audio from one channel of a CD drive was used as the subject matter.

Informal observations included comparison of the Convolvotron sound with Holophonic sound from a CD to subjectively understand and compare the techniques.

We took a digression into "how it works" to understand what we were working

with. This was a minor digression, however, and our main motivation was to try to use it "as is" in the interface rather than to examine the underlying methodology closely. The technology we used is briefly summarized here. The reader who is not interested in the technical details can skip to the next section.

Localized sound overview

The theme of localized sound research is that many humans can pretty well tell which direction sounds are coming from and that scientists ought to be able to figure out why and recreate the effect well enough to be able to make it happen in the laboratory. It seems to me that more progress is being made on the recreation than the figuring out why, but my review of the literature has not been extensive.

It is straightforward to explain how perceived amplitude and phase differences let us localize sound sources from side to side, i.e. if the sound gets to the left ear first and it is louder than what slightly later shows up at the right ear, the brain associates the source with our left side. How then are we able to identify that sounds are positioned above, below, in front or in back of us? The explanation goes on to describe the observation that the asymmetrically shaped outer ear or pinna seems to act on the sound which bounces around in it on its way to the eardrum. This modifies the spectral composition of the sound in a way which gives the brain more information to use in locating the source of the sound. The sound is modified differently depending on the elevation and azimuth of its source.

To recreate this in the laboratory, we could measure the effect of the pinnae on a known sound and simulate the effect electronically and see if the listener can tell the difference between the real and simulated sound. This is the basis of the technique we are using. The subject whose "ears" are to be measured sits in an anechoic (acoustically "dead") chamber with tiny microphones in both ear canals while a speaker is positioned at different locations in the room and a test sound is played and recorded via the ear microphones. The difference between the sound source, as played, and the recording made through the ear microphones is assumed to be the filtering contributed by the outer ears. The Convolvotron is then used to perform analogous filtering, in real time, on other sounds as they are delivered to the headphones of the listener, in hope that the filtering will provide the same localizing cues to the brain as the ear provides.

Note that the technique relies on a custom set of ear measurements for each observer and disregards many other aspects which might help the brain with sound localization such as reflection of sound from nearby walls, cognitive implications of the sound and familiarity with the sound, to mention a few.

We had two sets of "ear" data which were suitably modified to run four sound sources, Scott Foster's and the ears of a small female test subject. Given the assumption that few would go to the effort to get their ears measured for casual

listening and the fact that Dick Lyon is planning to do measurement of ears in the near future, we moved to the next phase, that of seeing if people could locate sound sources at all, using the "ears" we had.

The first user test

The goal of the first user test was twofold, to attempt to quantify (not rigorously) how a small set of people were able to localize sound using the techniques which we had available and to expose more of ATG to localized sound and get their feedback.

The experiment is described in appendix A and the conclusions from it are that subjects were very good at identifying left-right location but not very good at up-down or front-back location.

The "moving sounds" videotape

As a demonstration of using 3D sound to reinforce the notion that an observer was immersed in a 3D scene, we "produced" a videotape which demonstrated the use of localized sound as an environmental cue. This was done in anticipation of adding localized sound to the integrated space planning interactive demo and to capture localized sound and video on tape so that the effect could be observed without all the equipment needed to do it "live".

Laurie Vertelney made a Director prototype showing how a person's point of view might be controlled in the context of the space planning demo. This demo was an animation that showed how the view on the screen could be controlled by presenting the view from the viewpoints of different people within the scene. It was structured as several key, static scenes with motion transitions from one to the next and a narration which described the user interaction. We took the existing narration and added a background track of piano music (stolen) from a commercial CD. We positioned the sound sources spatially in the key frames so that the voice of the narrator would appear to be coming from the "person seated on the couch" and the background music would be coming from the television. We interpolated these positions during the visual transitions between scenes and initiated the transitions "by mouse" as the videotape played. The audio was recorded in stereo on the tape so that it could be played back through headphones.

When we played this tape to friends and test subjects later in the project, several things were almost universally observed. The first was that, if the tape was preceded by no discussion of the effect that we were trying to achieve, no one observed the effect. Some noted that the sounds seemed to be moving around, but the connection was not sufficiently obvious that a clear association was made between the sounds and the "sources". The second was that when the tape was played after the observer was told that "the narrator is sitting on the couch and the piano music is coming from the TV", the listener noticed the effect.

The implication of this experiment is that, for whatever reason, as we implemented it, localized sound did not add much to the demo for the untrained observer. It did not, however, rule out the value of localized sound after training, when the observer might have learned to expect the sound location to provide additional information. We did not investigate further the use of localized sound for the trained or the specialist, rather, we continued to look for common uses which might have wider appeal.

HIG working group feedback

The ATG Human Interface Group has weekly working team meetings which are used to show work in progress to the rest of the group and to get criticism and informal feedback. Because the localized sound demos rely on the user wearing headphones and having control of the interaction, the working group was accomplished as demos to groups of two. The goals were to let each member of the group hear the differences between just plain stereo, "Holophonic sound" and sound localized interactively by the Convolvotron. A demo showing how a user might access an audio source of information (the demo played a taped weather report) and have it move into auditory prominence was shown. The 3D videotape mentioned above was also played. Similar sessions were conducted for individuals working in sound from other parts of Apple to inform them and to solicit their opinions and impressions.

The responses to these sessions indicated that the localized sound technology, as we were using it, did not seem to produce a remarkable effect, but further, that the level of familiarity with and need for using *any sound* as part of the general Macintosh experience was very limited. This motivated a shift in focus from localized sound to an inquiry into how a Mac owner might use a machine to control his acoustic environment and what sort of things it might be able to provide in the future.

The "sonic configurator" user test

This led to a more formal user observation, this time with a wider audience of nine subjects selected from a pool of Apple employees maintained by the Human Interface Group. These experienced Mac users were asked to use a "Sonic Configurator" to position multiple sources of audio information so that they could control their acoustic environment while using their computer. The results of this observation are informal and indicate that there is interest in using the computer to control and access audio information much more than we currently do today.

The experiment also provided insight into some interface considerations which will need to be addressed to allow a user to configure and control a more complicated auditory environment. Discussions with the participants led to new insights as to how different sources of sound might be used in a personal computing

environment. The user test and results are described in more detail in appendix B.

Conclusions

Localized sound technology needs more work before it will be an important element of the personal computing environment.

The ability to localize sounds may be a desirable component of a rich audio environment for a personal computer, but it is not clear that the multi-channel, realtime 3D localization we tested is necessary or that its effect warrants the current cost of the capability. By continuing the basic research into sound localization which is planned in System Technologies, we stand to gain a firm basis in an area of technology which is young, imprecise and based on empirical techniques, some of which are entangled in the patent process. There are several people selling localized sound techniques and assistance to companies with lots of cash, but after this investigation, it is not clear that they can provide something which is fundamentally better than what we have looked at. As we understand a basis for localization and can produce better localized sounds, there are many ways in which we can introduce localized cues which have been "prerecorded" digitally and investigate the effect of these sounds in the interface. It is reasonable to assume sufficient DSP capacity today in the form of third party add-in boards to localize one channel, if needed. The hardware is probably ahead of the technology at this point.

The audio entertainment electronics business is changing rapidly as computers and digital signal processing chips are becoming common and we need to become actively involved in understanding the impact on our interfaces and our business.

The musician can buy very capable equipment which plugs and plays with other instruments and with personal computers. These give him both expanded performance and studio editing capabilities. It is reasonable to assume that the expectations of personal computer users will be affected by the emergence of this equipment. Sound synchronization and composition is an important part of multimedia document creation and editing and there are applications based on existing studio metaphors which we can learn from. It is interesting to note that the providers of instruments for entertainment audio are not the largest segment of the industry, compared to those who make and sell music to end users. In this context, we should think of our machines as delivery platforms for audio more than as tools for composition and editing by the small community of artists.

We need to start designing an integrated audio interface so that it's easy to hook up and use to control our auditory environment.

Audio is becoming recognized as an important aspect of personal computing. We have audio input as a standard feature on several new Macintosh models. One new

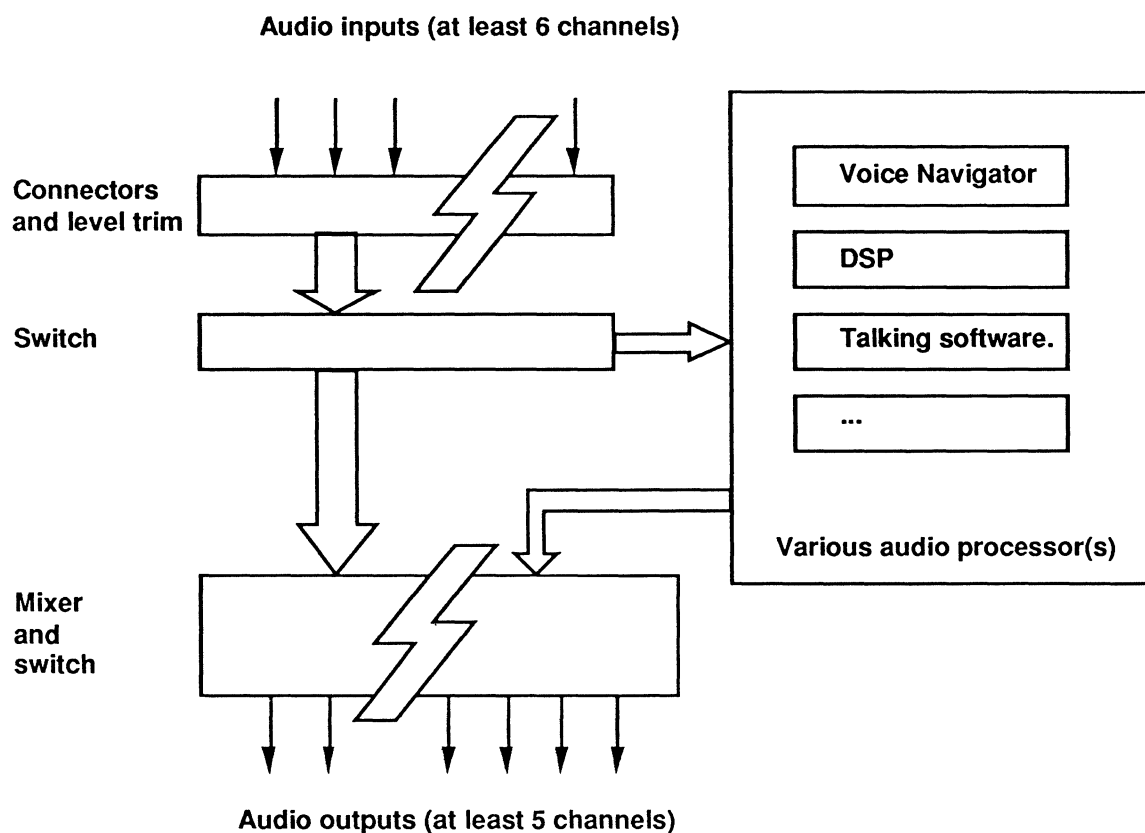
package design acknowledges the need for multiple audio interconnections and deals with the physical cabling interconnect with telephone, entertainment audio and video very elegantly. This addresses the physical hookup to the external audio world (which usually involves a rats nest of cables) quite nicely.

Once we have the ability to plug all the external connections in, what audio architecture supports control of all these channels? The ability to configure and to change configuration in response to different events would allow an end user to customize his audio environment and control otherwise distracting auditory events. Placement of sounds in a stereo field (with control of gain and pan) may be adequate for many users until better spatial techniques are developed. Given a basic mixing architecture, some sounds, like alert tones could be precomputed to appear in a given location and played into a mix of undiluted CD-quality personal entertainment audio (e.g. cheaper "Diskman" not CD-ROM). Similarly, as the technology is improved, DSP add-ons could inject "positionable" channels into the mix as desired. The end user could decide what equipment he wants to listen to and assemble it from a variety of components via familiar audio interfaces. Programmable mixing, signal conditioning and connector matching is the base technology for integration of sound sources under control of the personal computer. This would let the computer be more than just an additional source of sound for the user to deal with and it would position it as the master of all the other sources.

The mixer and the physical interconnections are the fundamental pieces for audio on a personal computer, not the DSP. The mixer is the setting for the DSP jewels which can do specialized audio effects. 44 Khz. audio from an entertainment disc shouldn't need to experience any loss in quality from digitization by going through a personal computer if only analog mixing is being done.

A block diagram for an audio capable personal computer of the future is included below, with the assumption that the switches and mixer are all software controlled and the amenities like digital, computer readable level meters are included. I have not specifically included converters for digital telephone interfaces, but have considered them only as audio inputs and outputs. This is clearly an oversimplification, as are the details of the "Audio processors" box, which is left vague to allow inclusion of any of the audio boxes which are becoming available.

This audio configuration is easy to prototype with a blob of wires and boxes today and experimentation with the interface can be done in parallel with DSP development.



Basic architecture for personal computer audio

The Sonic Configurator experiment indicated that once all these sources of audio are physically attached to the computer, setting them up so that the levels are comfortable for the individual and so that the sound sources work together is an interesting interface problem which doesn't have much precedent in our everyday lives.

Our prototype took the obvious "graphical" approach to positioning sound sources, but the experiment uncovered some problems with this familiar approach. The visual representation which was used confuses effective sound placement and tends to dominate it. For example, one subject adjusted the position of a sound with closed eyes and upon opening them, observed that the visual representation was not symmetric and changed it, making the assumption that visual symmetry and auditory symmetry were somehow related. One subject drew the analogy to the "graphic equalizer" syndrome in which one buys a new device for the stereo, sets it up once to compensate for some subjective properties of his sound environment and ignores it thereafter.

An interface for critical audio control would allow adjustment of the sound with

ones eyes closed, allowing maximum concentration on what one hears by shutting down the visual channel. This raises questions of what (non-visual) techniques are appropriate to select and highlight a sound for further manipulation? What controls should then be applied to the selected sound? Volume and pan between left and right ears, at least. What 3D abstraction of the controls would allow stereo placement now and more general spatial placement within the same metaphor when the technology becomes better understood? What physical controls would make this easier; buttons, dials or rollers?

Being able to control interruptions provides new things to set up beyond the current speaker volume slider. Given an audio environment which includes external event notification, in addition to selecting the auditory icon, the user should be able to determine how the event will be presented e.g. as a high priority override, as an informational item or as an event to be explicitly ignored. One can imagine assigning the telephone bell to different priorities depending on whether one was expecting a call-back, doing light activities like reading mail or attempting intense concentration. Providing and managing a more complex audio context is a consequence of hooking up to many external sources.

One could imagine set of audio effects comparable to the HyperCard visual effects - fades, dissolves, etc. so that developers can do whizzy, comfortable and transparent things for the end users. For example, is applying "warble" to an audio source perceptually equivalent to the "marching ants" used to show selection visually?

Finally, but perhaps most importantly, any audio accessory from Apple has to be QUIET when no signal is going through it, no matter how inexpensively it can be made. If our experience with the Convolvotron is typical, electrical noise is a problem for audio equipment in the computer environment and care has to be taken to keep audio noise out.

There are several small projects which we can learn from soon.

Several of the subjects observed that control of multiple audio channels allowed access to a high bandwidth sensory mode which we do not commonly use in our computing environment, but which we frequently and naturally use on a daily basis. Our ability to simultaneously monitor multiple conversations and "pick out" the one which is currently of interest to us (the cocktail party effect) is well developed. Being able to control how multiple audio sources are presented to us would allow establishing a "Locus of Control" within which we could use our mental ability to monitor multiple auditory threads and switch between them.

An experiment would be to determine the effectiveness of using a videotape with multichannel audio as an educational tool. An example might be a history lesson with commentaries by multiple "guides" (as shown in Tim Orens Grolier project) which could be presented simultaneously and "tuned in" by the user, both by

controlling volume or location of the guides voice and by "mental switching" between them. A similar, but realtime, equivalent would be a setup which allowed monitoring multiple audio channels of broadcast media during the impending Iraq crisis.

Another interesting experiment would evaluate how using your computer to control audible interruptions in an "open office environment" affected the ability to concentrate and perform creative work. It would include observations of one of us "knowledge workers" in an open office environment, comparing concentration and span of attention as things are today, as they would be with personal audio and headphones as a shield against interruption and as they would be with a personal computer controlling the audio environment via ambient sound blocking headphones. I think this experiment is interesting because of the large body of users (at least locally) who use Macs in an open environment and because of the interest expressed by the people who participated in the user tests.

I believe that either experiment can be prototyped with an off-the-shelf MIDI controlled mixer and a "second prototype interface" piece of software yet to be written.

Appendix A: "The first experiment"

The experiment which we embarked upon initially evolved over time, as shown in the following matrix of test subject cards vs. conditions which we varied as we learned more.

All tests were based on the user answering questions about the location of sounds which were played at 32 different locations in the auditory field. The locations were all combinations of 8 azimuth (-180, -150, -90, 30, 0, 30, 90, 150) and 4 elevation values (-36, 0, 18, 54), selected so as to be subjectively categorized by responses to the questions "front, middle or back", "left, middle or right" and "up, middle or down". The tones were presented to the subject in a random order. As a tone was played, only one of the questions was asked, so each tone was played three times to determine the perceived location.

These locations were selected so that they coincided with locations at which Head Related Transfer Function (HRTF) data was measured. This eliminated the effect of interpolation in the Convolvotron which extends the resolution so that sounds can apparently be positioned anywhere in the auditory field. This interpolation seems to work, but it is yet another source of uncertainty which could be easily avoided by careful location selection. Only 32 locations were used to keep the tests short (but the test still took about 20 minutes).

The initial tests were performed using the HRTF localization coefficients of the subject named SF, but "whitened" by Dick Lyon to remove the effect of resonances which seemed to be artifacts of the measurement process (these are indicated by "Mod SF" in the following table). Subsequently, we also used the SF and SDO HRTF coefficients which were supplied with the Convolvotron.

For all subjects the test used a repeated, percussive tone (Shaker) and later the test length was doubled to include a more musical note with different frequency and subjective characteristics (Kalimba).

Finally, we changed the test to make the responses simpler and less ambiguous by forcing a choice to "up or down", "left or right" or "front or back" by eliminating the "middle" response which seemed to not clearly convey the axis being tested.

Card #	Date	Mod	SF	SF	SDO	Shaker	Kalimba	No	Middle
--------	------	-----	----	----	-----	--------	---------	----	--------

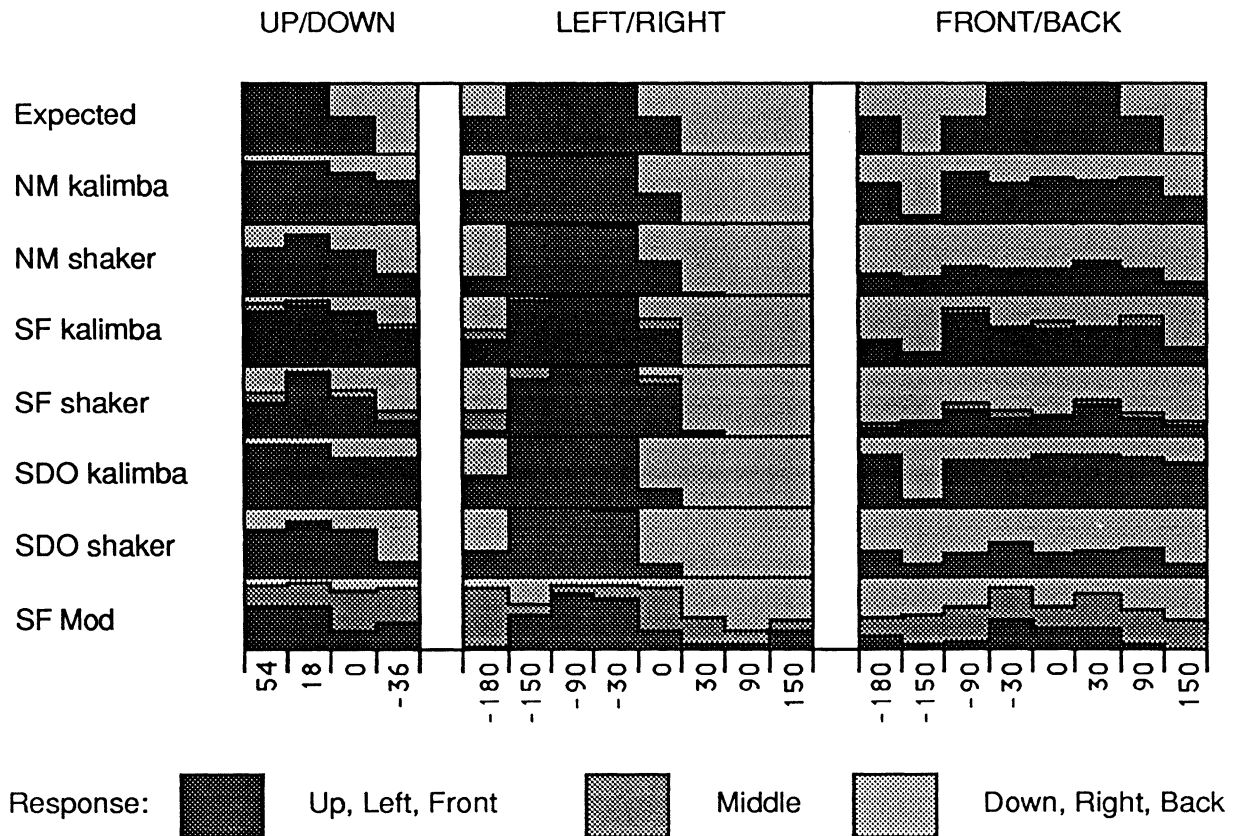
21	09-26-90	✓				✓			
20	09-26-90	✓				✓			
19	09-26-90	✓				✓			
18	09-26-90	✓				✓			
17	09-26-90	✓				✓			
16	09-27-90	✓				✓			
15	09-28-90		✓			✓	✓		
14	09-28-90		✓			✓	✓		
13	10-01-90		✓			✓	✓		
11	10-01-90		✓			✓	✓		✓
10	10-01-90		✓			✓	✓		✓
9	10-01-90		✓			✓	✓		✓
8	10-01-90		✓			✓	✓		✓
7	10-05-90			✓		✓	✓		✓
6	10-05-90			✓		✓	✓		✓
5	10-05-90			✓		✓	✓		✓
4	10-05-90			✓		✓	✓		✓
3	10-08-90		✓			✓	✓		✓
2	10-08-90			✓		✓	✓		✓

Subject test card vs. test conditions

The data was informally analyzed by looking at summations of responses on a per-subject basis in three ways: up/down responses were compared to elevation, left/right responses were compared with azimuth and front/back responses were compared with azimuth.

A more rigorous graphical comparison is shown below, with each row characterizing a subset selection from the responses of all subjects. The top row of the graph shows the ideal or "expected" response.

Subsequent review of the literature disclosed that several of the sound locations (those with azimuth of -150 and 150 degrees and elevation of 54 and 18 degrees) had been selected in regions which experiments had found to be hard for subjects to localize accurately (mean positional error > 20 degrees) so summations were recalculated to omit these points from the analysis. Significant differences were not noted when these locations were omitted from the summary of the data. The graph below omits the locations in question.



Summary of user responses vs. point elevations and azimuths

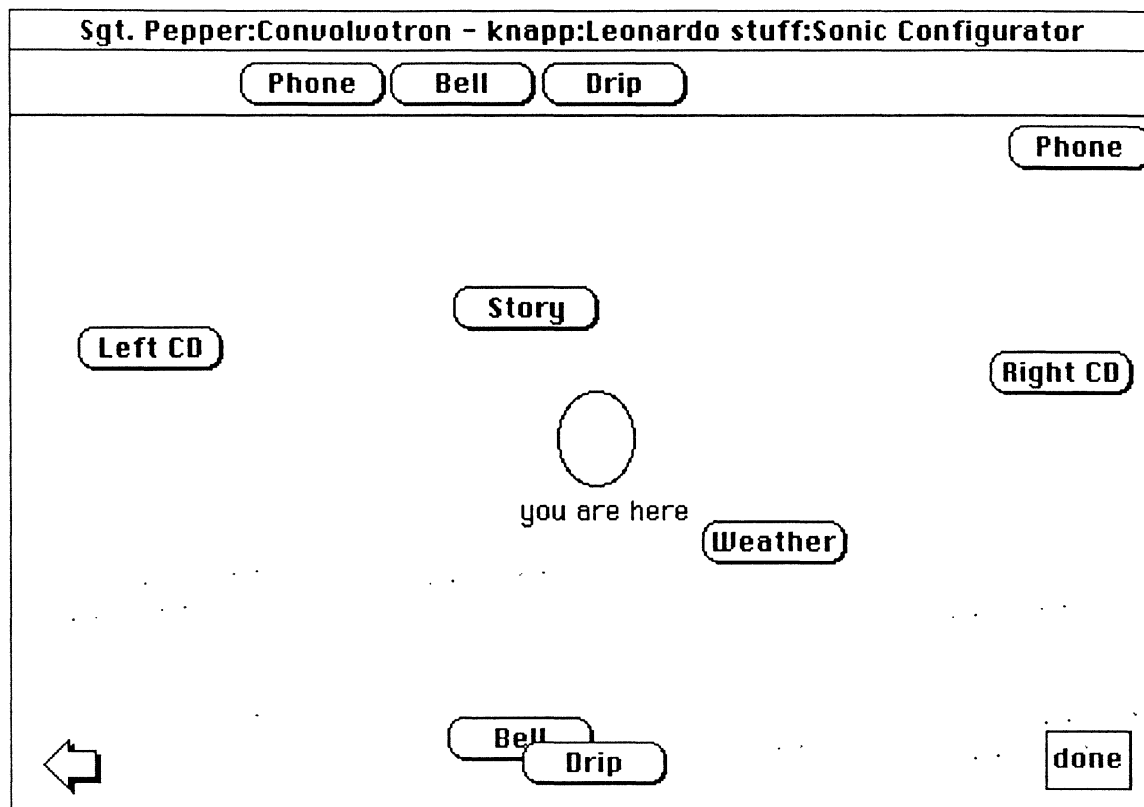
The conclusion is that there was a visible consistency among the left-right responses and the up-down and front-back responses were much less like the "expected" responses and less similar to each other. This motivated investigating using just stereo to position some sound sources and a look at the more general problem of controlling multiple sound channels with a personal computer.

Appendix B: The "Sonic Configurator" User Test

The presumption of this "user exposure" was that the user was trying out a new computer which had more control of sounds and audio sources in the interface than are common today. The test was based on a prototype "Sonic Configurator", a program which allowed the user to set up the audio environment which was to be controlled by the computer. It assumed connection to numerous external sound sources including a stereo audio source of "personal entertainment", a (presumably current) weather report, a voice track of information and/or entertainment ("Coyote Stories", as told by an Indian storyteller) and some sources of "alert" sounds which could be used to notify the user of events. The alert sounds included a large bell, a repeated dripping sound and the ubiquitous telephone. The interface could be configured by dragging a button associated with the sound source around on a card and listening to the effect of positioning the sounds via stereo headphones. Volume control and balance were adjusted using physical sliders on a mixer near the keyboard, avoiding the interesting question of how to simultaneously set position in the auditory field and volume of a source. The user's actions were recorded by an over-the-shoulder view of the screen including stereo audio of what the observer was hearing during the positioning of the buttons. The final configurations reached were retained in the HyperCard stack.

The users were drawn from within Apple. The selection criteria was "experienced Mac users" and if they used headphones while they worked, so much the better. It was made clear that it would be an audio test and that if they had known hearing loss, they might not get as much out of the test. Users included programmers and writers, people who must typically work without interruption and concentrate during some aspects of their work.

In preparation for the test, the subject chose a CD as a source of personal entertainment music. The test proceeded in several stages; the first card allowed positioning of the two stereo channels of personal entertainment music. The second card added two voice sources; the weather report and the Coyote stories. The goal was to place the sounds so that it was easy to discriminate between them while they were played simultaneously. The third card added the ability to place three alert sounds in the auditory field, the bell, the drip and the phone. (Only four sounds were simultaneously active at any one time.) Finally, a demonstration of how sounds might be controlled collectively as a group was presented. It included clicking on an icon to access the weather report and bring it to the center of focus, while reducing the volume of the personal entertainment sounds. Another example involved initiating an "alert sequence" which caused the large bell sound to become progressively more prominent by changing its volume, position and the volume of other sounds in the scene.



Control window of "Sonic Configurator"

Placement of the sounds was done by positioning buttons on a card on the screen. The initial position of the buttons was above a line drawn across the top of the card. Some buttons (Phone, Bell and Drip) "renewed" themselves after they were positioned below the line so that multiple copies of that sound could be positioned. The users position in the visual representation was indicated by an oval labeled "you are here", but no other detail was added to imply what actual space the sounds were being located in and the instructions were to place the sounds based on what sounded good to the user. The location of the button on the card was actually used to specify Convolvotron coordinates. The horizontal dimension mapped the left edge of the card to -180 degrees (directly behind the head), the center of the card to 0 degrees (directly in front) and the right edge on around to 180 degrees (behind the head again). The vertical extent of the card was mapped to the elevation of the sound source, the bottom of the edge mapped to -36 degrees from the horizontal and the top to +54 degrees. Noone was overly concerned about the ambiguity of where the sounds "really" were being placed or what the location of the button on the card meant.

In the phase where just the stereo sources were placed, users observed the left and right channel conventions, but otherwise there was variation in both position (some high, some low, some front, some back) and symmetry (6 symmetric, 3 not so symmetric) of the resulting sound locations.

In the second phase, the goal was to position two voice sources in addition to the stereo so that all could be heard clearly and the user could (mentally) switch between different sources. Examining the visual position of the buttons over the nine users indicated a wide variety, with no apparent patterns or similarities between the users, except the symmetry between the left and right CD channels observed in the first phase.

In the third phase, the audio environment was a bit more cluttered. The buttons were typically distributed well around the screen, rather than clustered or overlapped with each other, but again, no clear patterns or similar configurations were observed.

The fourth phase was a demonstration to show the participant how sounds might be choreographed together to provide controllable notification of an event or to make access to an external audio source easy, without having to remove headphones. This phase was intended to initiate discussion of possible uses for a controllable sound environment.

Users' impressions

The gathered from discussion with the participants included comments on their physical environment, possible uses for audio effects, subjective reactions to localized sound and comments on the interface to such an audio environment.

Environment

Subjects compared the way they use their Macintoshes now with how they might use them if they had richer audio environments. Most subjects worked in a cube environment and the time spent in front of the Macintosh ranged from 1/3 of the day to most of the day. Several subjects wear headphones while using their computers, but most did not. Some liked the headphones, which blocked ambient room noise (key clicks from other computers, noisy fans, people talking in next cube, others' phones, were all mentioned as sources of distraction). One subject who doesn't use headphones now did not know what subject matter would help concentration and suggested classical tapes or natural sounds.

The subjects who use headphones today use them to help maintain their concentration by providing background audio to mask other distractions and to keep people from interrupting them. Some even admitted to using headphones with no music as a social "mask" to keep from being bothered. These people suggested that the ability to detect when someone stood in their doorway would be a welcome capability because using the headphones to screen local sound lets people sneak up on you. The ability to control how much ambient sound would be audible is desirable, rather than just eliminating it altogether. One subject suggested the

computer provide a "door squeak" to let you know someone is entering your cube. One subject who already uses an audio mixer to combine entertainment sound and computer sounds noted that the telephone was next.

Noone appeared uncomfortable with the notion of having their computer control their audio environment, although some, whose jobs required them to be "interruptible" by people in their group felt that they would not feel comfortable using headphones because of the social barrier it presents to others. In general, the level of enthusiasm about controlling interruptions was quite high.

One person who moves between his office and lab felt that headphones with a cord would be too restrictive but that cordless headphones would work well.

Uses

In addition to masking unwanted noise, a range of uses of sound were suggested, from alarms or meeting notifications (the user's watch beeped once during the session) to alerts to signal completion of long processes (compiles, table of contents generation). There was not as much enthusiasm or as many specific suggestions in this area because computers beep when they should already and no really novel uses emerged.

One programmer who uses several computers simultaneously felt that audio would be an easy means to let him switch between the several machines which he uses and integration of localized sounds into a set of headphones would make it easy to tell which needed his attention.

Uses in the home included blocking out the TV in the next room, providing timers while cooking, audio mixing of the "baby monitor" and doorbell.

One person articulated the best "use" of an integrated audio interface (as opposed to "sounds in the interface") as that of being able to control the users focus. By being able to exclude some sounds (irrelevant conversations, etc.) and include others (newscast in the background or telephone call) when we would like to, we have the ability to prioritize and control interruptions and monitor multiple sources of information using whatever cognitive mechanism manages the "cocktail party" effect for us.

Localized sound

When specifically asked, some did notice a slight expansion of the auditory space, e.g. the sound moving "out of head" vs. just along a line between the ears, but an A/B comparison between audio directly from the CD player and the same audio localized through the Convolvotron made it readily apparent that there was frequency-specific attenuation happening and the effect on the perceived audio

quality was unacceptable. (Bass attenuation was very noticeable.)

Reactions to the moving alert sounds indicated that it was obvious when they moved from the side to the center, but the vertical and front/back components of movement were not usually mentioned.

Interface

Subjects generally liked to be able to position sounds. The ability to allow for things like "my telephone ear" seemed convenient. One subject wanted volume control via the computer rather than external sliders on the mixer.

The graphical interface used in the experiment drew mixed reviews. One subject noted that the present interface in the control panel, the Speaker Volume slider, is clear and simple, in contrast to the multiple source version presented in the experiment. It was noted that "up and down movement of button didn't do much" and in at least one case, it was assumed to be the volume control associated with the sound. One subject related the interface for sound positioning to the "Graphic Equalizer Phenomenon" seen in home stereo equipment wherein the purchaser buys a new stereo component, takes great pains to set it up once and obtain subjectively perfect results and thereafter never touches the controls again.

References

Oldfield, Simon R. and Parker, Simon P. A, "Acuity of sound Localization: a topography of auditory space. I. Normal hearing conditions", *Perception*, 1984 volume 13, pp. 581-600

Wenzel, Elizabeth M., Wightman, Frederic L. and Foster, Scott H., "A virtual display system for conveying three-dimensional acoustic information", *Proceedings of the Human Factors Society, 32nd Annual Meeting*, 1988

Singer, David, "Bibliography of Localized Sound Papers", Apple ATG, Informal communication.

Wenzel, Elizabeth M., Wightman, Frederic L., Foster, Scott H. and Kistler, Doris J., "Realtime Synthesis of Localized Auditory Cues", Presented at CHI '89 Austin Texas, April 30 - May 5, 1989